AMENDMENTS TO THE CLAIMS:

1-30. (Canceled)

31. (Currently amended) A system for identifying genes, comprising:

a pattern database comprising patterns of amino acids;

an input device for inputting a genomic DNA sequence; and

a processor which is configured to:

translate an open reading frame (ORF) of said DNA sequence into an amino acid translation;

assign weights, w_i, to said patterns of amino acids, said weights being given by the equation $w_i = \log p_i - \log q_i$, where p_i is a probability that a pattern[[, t_i,]] matches an actual amino acid sequence at a fixed location, and q_i is a probability that said pattern[[, t_i,]] matches an amino acid translation of a non-coding ORF;

locate in said amino acid translation occurrences of said weighted patterns, and assign a coding quality measure for said ORF which is given by $Ws = \sum_{i=1}^{m} w_i$, where m is the number of ~~based on a sum of said weighted~~ patterns which are located in said amino acid translation of said ORF ; and

identify said open reading frame as including a putative gene if a value of said coding quality measure is greater than a predetermined threshold value .

32. (Previously presented) The system according to claim 31, wherein said processor translates a plurality of open reading frames in said DNA sequence into amino acid translations, and locates in each amino acid translation occurrences of said patterns to determine whether each said plurality open reading frames includes a putative gene.

33. (Previously presented) The system according to claim 32, wherein said patterns comprise biologically significant patterns of amino acids in amino acid sequences.

34. (Previously presented) The system according to claim 31, wherein said processor identifies a match of a pattern from said pattern database in said amino acid translation.

35. (Previously presented) The system according to claim 34, wherein said patterns are derived from a parent database comprising at least one amino acid sequence.

36. (Previously presented) The system according to claim 34, wherein said patterns are derived from a parent database comprising at least one amino acid sequence fragment.

37. (Previously presented) The system according to claim 34, wherein said patterns are derived by using a pattern discovery algorithm.

38. (Previously presented) The system according to claim 34, wherein said patterns are derived by using the Teiresias algorithm.

39. (Previously presented) The system according to claim 34, wherein said ORF comprises a portion of said DNA sequence between a start codon and a stop codon.

40. (Previously presented) The system according to claim 34, wherein said processor reports said ORF as a putative gene when a predetermined number of pattern matches is identified in said amino acid translation.

41. (Previously presented) The system according to claim 34, wherein each pattern is assigned a weight depending upon a relevance of said pattern in determining whether said ORF comprises a putative gene.

42. (Previously presented) The system according to claim 34, wherein said processor is

configured to select a start codon which results in a greatest value of said coding quality measure, in a case in which plural start codons match the same stop codon.

43. (Previously presented) The system according to claim 34, wherein said match is identified using a predetermined pattern matching algorithm.

44. (Previously presented) The system according to claim 34, further comprising:

a memory device for storing data and instructions to be executed by said processor.

45. (Previously presented) The system according to claim 34, further comprising:

a display device for displaying an output from said processor.

46. (Currently amended) A method of identifying genes, comprising:

providing a pattern database comprising patterns of amino acids;

determining an open reading frame (ORF) in a genomic DNA sequence;

generating an amino acid translation for said ORF;

assigning weights, w_i, to said patterns of amino acids, said weights being given by the equation $w_i = \log p_i - \log q_i$, where p_i is a probability that a pattern[[, t_i,]] matches an actual amino acid sequence at a fixed location, and q_i is a probability that said pattern[[, t_i,]] matches an amino acid translation of a non-coding ORF;

locating a match of said weighted patterns from said pattern database, in said amino acid translation and assigning a coding quality measure for said ORF <u>which is given by $Ws = \sum_{i=1}^{m} w_i$,</u>

<u>where m is the number of</u> ~~based on a sum of said weighted~~ patterns which are located in said amino acid translation of said ORF;

identifying said ORF as including a putative gene if a value of said coding quality measure is greater than a predetermined threshold value; and

displaying a result of said identifying said ORF as including a putative gene.

47. (Previously presented) The method according to claim 46, wherein said pattern database is generated from a database comprising at least one amino acid sequence.

48. (Previously presented) The method according to claim 46, wherein said pattern database is generated from a database comprising at least one amino acid sequence fragment.

49. (Previously presented) The method according to claim 46, wherein said probability, p_i , is calculated based on a training set .

50. (Previously presented) The method according to claim 49 , wherein said probability, q_i, is calculated by computing a number of occurrences of a pattern in ORFs that are not identified as coding in said training set.

51. (Previously presented) The method according to claim 46, further comprising:

displaying said match of said pattern in said amino acid translation.

52. (Previously presented) The method according to claim 46, wherein said pattern database is generated using the Teiresias algorithm to derive said patterns from a parent database.

53. (Currently amended) A programmable storage medium tangibly embodying a program of machine-readable instructions executable by a digital processing apparatus to perform a method for identifying genes, said method comprising:

providing a pattern database comprising patterns of amino acids;

determining an open reading frames (ORF) in a given genomic DNA sequence;

generating an amino acid translation for each ORF;

assigning weights, w_i, to said patterns of amino acids, said weights being given by the equation $w_i = \log p_i - \log q_i$, where p_i is a probability that a pattern[[, t_i,]] matches an actual amino acid sequence at a fixed location, and q_i is a probability that said pattern[[, t_i,]] matches an amino acid

translation of a non-coding ORF;

locating a match of said weighted patterns from said pattern database, in said amino acid translation and assigning a coding quality measure for said ORF which is given by $Ws = \sum_{i=1}^{m} w_i$, where *m* is the number of ~~based on a sum of said weighted~~ patterns which are located in said amino acid translation of said ORF;

identifying said ORF as including a putative gene if a value of said coding quality measure is greater than a predetermined threshold value ; and

displaying a result of said identifying said ORF as including a putative gene.

54. (Previously presented) The system according to claim 33, wherein said processor determines for each pattern in said pattern database whether the pattern is present in said amino acid translation by locating instances of said patterns in said amino acid translation, until a sum of weights corresponding to all patterns with matches in said amino acid translation exceeds a predetermined threshold, at which point said processor identifies said ORF as a putative gene.

55. (Previously presented) The system according to claim 31, further comprising:

a parent database comprising a plurality of amino acid sequences, said patterns in said pattern database being derived from said plurality of amino acid sequences by using a pattern discovery algorithm;

a memory device for storing data and instructions to be executed by said processor; and

a display device for displaying an output from said processor.

56. (Previously presented) The system according to claim 55, wherein said open reading frame (ORF) comprises a portion of said DNA sequence between a start codon and a stop codon,

wherein said processor identifies a match of a pattern from said pattern database in said amino acid translation by using a predetermined pattern matching algorithm,

wherein each pattern is assigned a weight depending upon a relevance of said pattern in

determining whether said ORF comprises a putative gene, and

wherein said ORF is reported as a putative gene when either a predetermined number of pattern matches is identified in said amino acid translation, or a sum of weights corresponding to all patterns with matches in said amino acid translation exceeds a predetermined threshold.

57. (Previously presented) The system according to claim 31, wherein said processor accesses said pattern database to retrieve said patterns from said pattern database.

58. (Previously presented) The system according to claim 31, wherein said processor is electrically coupled to said input device and said pattern database.

59. (Currently amended) A system for identifying genes, comprising:

an input device which inputs a genomic DNA sequence; and

a processor which is configured to:

access a pattern database comprising a plurality of patterns of amino acids;

translate an open reading frame (ORF) of said DNA sequence into an amino acid translation;

assign weights, w_i, to said patterns of amino acids, said weights being given by the equation $w_i = \log p_i - \log q_i$, where p_i is a probability that a pattern[[, t_i,]] matches an actual amino acid sequence at a fixed location, and q_i is a probability that said pattern[[, t_i,]] matches an amino acid translation of a non-coding ORF;

locate in said amino acid translation occurrences of said weighted patterns, and assign a coding quality measure for said ORF which is given by $Ws = \sum_{i=1}^{m} w_i$, where m is the number of ~~based on a sum of said weighted~~ patterns which are located in said amino acid translation of said ORF ; and

identify said open reading frame as including a putative gene if a value of said coding quality measure is greater than a predetermined threshold value.

60. (Currently amended) A system for identifying genes, comprising:

an input device which inputs a query genomic DNA sequence;

a processor which is configured to:

access a pattern database comprising a plurality of patterns of amino acids;

translate an open reading frame (ORF) of said DNA sequence into an amino acid translation;

assign weights, w_i, to said patterns of amino acids, said weights being given by the equation $w_i = \log p_i - \log q_i$, where p_i is a probability that a pattern[[, t_i,]] matches an actual amino acid sequence at a fixed location, and q_i is a probability that said pattern[[, t_i,]] matches an amino acid translation of a non-coding ORF;

locate in said amino acid translation occurrences of said weighted patterns, and assign a coding quality measure for said ORF which is given by $Ws = \sum_{i=1}^{m} w_i$, where m is the number of ~~based on a sum of said weighted~~ patterns which are located in said amino acid translation of said ORF; and

identify said open reading frame as including a putative gene if a value of said coding quality measure is greater than a predetermined threshold value ;

a display device for displaying an output of said processor, said output including an occurrence of said patterns in said amino acid translation,

wherein said patterns comprises patterns derived using a Teiresias algorithm,

wherein said open reading frame (ORF) comprises a portion of said DNA sequence between a start codon and a stop codon, and

wherein said processor identifies a match of a pattern from said pattern database in said amino acid translation by using a predetermined pattern matching algorithm.